Journal of Medical Virology 49:155–160 (1996)

Discrepancy of Hepatitis C Virus Genotypes as Determined by Phylogenetic Analysis of Partial NS5 and Core Sequences

Zhibing Yun, Claudia Lara, Bo Johansson, Ivette Lorenzana de Rivera, and Anders Sönnerborg

Division of Clinical Virology, Institution of Immunology, Microbiology, Pathology and Infectious Diseases, Karolinska Institutet, Huddinge University Hospital, Huddinge, Sweden (Z.Y., C.L., B.J., A.S.); Department of Microbiology, Virology Unit, Universidad Nacional Autonoma de Honduras, Honduras (C.L., I.L.d.R.)

The use of phylogenetic analyses of partial NS5 and core regions for hepatitis C virus (HCV) genotyping was evaluated by analysing seven Honduran and 24 European HCV strains. Core primers were designed with which HCV genotypes 1, 2, and 3 were readily amplified. The reliability of phylogenetic analysis of a 111-bp core sequence was verified by comparing the typing results with those obtained using the whole core gene of 52 reference strains. Accordant genotypes (1a, 1b, 2b, and 3a) were obtained when phylogenetic analyses were undertaken on both the partial core and a 222-bp NS5 sequence in all of the European HCV strains. Genotypes 1a, 1b, and 3a were identified among the Honduran strains by phylogenetic analysis of the partial NS5 sequence. Interestingly, two of three Honduran type 3a strains, as determined by the NS5 sequence analysis, turned out to be type 1a by core sequence analysis. These two strains were also classified as type 1a, but not 3a, by a core type-specific PCR. Furthermore, the E2/NS1 regions were similar to HCV-PT, a representative strain of genotype 1a. The results indicate that chimeral HCV strains exist, although in most cases a good concordance is found when phylogenetic analysis of partial core and NS5 sequences are used for genotyping. This finding should be taken into account when HCV is genotyped by phylogenetic analysis of a partial HCV sequence from a single genomic region.

KEY WORDS: HCV, NS5, core, Honduras

INTRODUCTION

Hepatitis C virus (HCV) is the major cause of blood-borne non-A, non-B hepatitis [Houghton et al., 1991]. It is distributed widely in the world [Bukh et al., 1993; McOmish et al., 1994]. Analysis of the HCV genotypes has become the major tool for HCV classification and characterization [Simmonds, 1995]. Since various genotyping systems have been used [Nakao et al., 1991; Okamoto et al., 1992a; Simmonds et al., 1993; Stuyver et al., 1993], comparison of genotype data is difficult. At present, phylogenetic analysis of sequence data is considered to be the most reliable method and is used as reference for the evaluation of other genotyping assays [Simmonds, 1995; Lau et al., 1995]. Phylogenetic analysis of a partial NS5 sequence [Simmonds et al., 1993] or the entire core and E1 genes [Bukh et al., 1993; Bukh et al., 1994] has so far been successfully used. When comparing a number of NS5, E1, and core sequences, no evidence of recombination of HCV strains was found [Simmonds et al., 1994]. However, it has been stated that further studies are needed to clarify this issue [Simmonds, 1995]. In the present study, we evaluated the use of phylogenetic analyses of several genomic regions for the characterization of HCV strains obtained from Honduras, a country from which no information regarding HCV genotypes has been reported previously.

MATERIALS AND METHODS

Patients

Sera were obtained from eight anti-HCV positive patients in Honduras. Three patients were polytransfused (HO 2, HO 3, HO 4), two patients were detected in connection with blood donation (HO 1, HO 7), and three were prison inmates (HO 5, HO 6, HO 8). Twenty-four HCV strains from Sweden and Norway were also analysed together with the Honduran strains.

Reverse Transcription and PCR

HCV RNA was extracted from 100 μl serum [Yun et al., 1993] and subjected to a combined RT-PCR, using

Accepted for publication January 10, 1996.

Address reprint requests to Anders Sönnerborg, Div. of Clinical Virology, Inst. of Immunology, Microbiology, Pathology and Infectious Diseases, Karolinska Institutet, Huddinge University Hospital, S-141 86 Huddinge, Sweden.

Sequence data from this article have been deposited with GeneBank under accession numbers U28019–U28037.

TABLE I. Sequences of Primers Used for Amplification of the HCV Core and E2/NS1 Regions

Name	Region	Position of 5′ base[a]	Polarity	Sequences (5′ to 3′)[b]
COU	Core	82	+	GGICAGATCGTTGGTGGAGT
COD	Core	413	−	GGGATGTACCCCATGAGGTC
CIU	Core	119	+	GGGGCCCIIGITTGGGTGTG
CID	Core	383	−	CACGTIAGGGTATCGATGAC
YU1	E2/NS1	1,350	+	CT(G/A)CTCCGGATCCCACAAGC
YU2	E2/NS1	1,426	+	(T/G)(C/T)TCCATGGTGGGGAACTGG
YD1	E2/NS1	1,614	−	CATTGCAGTTCAGGGCCGT

[a]The position of nucleic acid is numbered, as earlier described [Choo et al, 1991].
[b]I, inosine.

5 units MuLV reverse transcriptase (Boehringer Mannheim, Germany), 10 units RNasin (Boehringer Mannheim, Germany), 1 unit *Taq* polymerase, 0.2 mM of each nucleotide, and 20 pmol of outer PCR primers. The RT-PCR program consisted of incubation at 42°C for 45 min and heating at 95°C for 3 min, followed by denaturation at 95°C for 1 min, annealing at 45°C for 2 min, and elongation at 72°C for 3 min, in total 35 cycles. The first round PCR product was diluted to 10^{-3} to 10^{-6} before being subjected to further amplification. The nested PCR contained the same components as the first PCR, except for the reverse transcriptase and RNasin. The amount of inner primers was 5 pmol. The nested PCR programmes were denaturation at 95°C for 30 sec, annealing at 55°C (for amplification of core and E2/NS1 regions) or 50°C (for amplification of NS5 region) for 30 sec, elongation at 72°C for 1 min, in total 30 cycles. The primers used for amplification of the NS5 region have been described earlier [Enomoto et al., 1990; Simmonds et al., 1993]. The primers designed for amplification of the core and E2/NS1 regions are shown in Table I. One of the nested primers in each pair was biotinylated, and the other one was coupled with M13 universal primer (CGACGTTG-TAAAACGACGGCCAGT) at the 5′ end of the primer.

Direct Sequencing

The biotinylated PCR product was directly immobilized onto streptavidin-coated magnetic beads (Dynal AS, Oslo, Norway). In some cases, the PCR products were purified using a MicroSpin column™ (Pharmacia Biotech, Uppsala, Sweden) before immobilization. Single-stranded DNA (ssDNA) was generated by denaturation with 0.1 M NaOH. The sequence was carried out with FITC-labeled M13 universal primer using an automated laser fluorescent sequencer (A.L.F., Pharmacia Biotech, Uppsala, Sweden).

Phylogenetic Analysis

Nucleotide sequences were aligned using the Pileup program of the GCG Wisconsin Sequence Analysis Package (version 8). The alignments were converted to Phylip (version 3.5) input format. Sequence distances were calculated using the DNADIST program (maximum likelihood setting). Clustering of the distance data and assembly of the trees were carried out with the NEIGHBOR program. The robustness of the trees was estimated by bootstrap analysis of the aligned sequence data using SEQBOOT, followed by the DNADIST and NEIGHBOR programs on the 100 trees produced. The tree file obtained was fed into the CONSENSE program to produce a consensus tree. The number of times a specific cluster of strains (a subtype) appeared among 100 trees was obtained from the output file.

RESULTS

Analysis of NS5 Sequences

The NS5 sequences were amplified from seven of eight Honduran patients. Phylogenetic analysis was carried out on a 222-bp sequence, as earlier described [Simmonds et al., 1993]. When the Honduran strains were compared with the genotypes from 43 published strains [Simmonds et al., 1993], the strains HO 1, 6, and 7 were classified as type 1a, HO 5 as type 1b, and HO 2, 3, and 4 as type 3a (Fig. 1). The bootstrap data (100 trees) for the clustered subtypes were 100 for 1a, 1b, 2a, 2b, 3a, 3b, respectively, and 99 for 2c. This data showed the high degree of robustness of genotyping by the phylogenetic analysis of the selected NS5 sequence.

Analysis of Core Sequences

Based on published core sequences [Bukh et al., 1994], we designed two pairs of primers (Table I), which were expected to anneal readily to the sequences of genotypes 1, 2, and 3, respectively. One of the type 3a samples, as determined by NS5 sequencing, was not tested due to lack of serum (HO 2). All of the other six sera were amplifiable using our primers. As suggested by the reported sequences [Bukh et al., 1994], a 111-bp sequence located at the 5′ end of the amplicon should contain enough genetic information to be able to distinguish the existing genotypes. Therefore, we decided to construct phylogenetic trees on this 111-bp sequence. To verify the reliability of phylogenetic analysis of the selected region for genotyping, this region was analysed and compared with previous typing results using the whole core gene (573 bp) from 52 published strains [Bukh et al., 1994]. The genotyping results were identical. The bootstrap data (100 trees) for the clustered subtypes using 111-bp sequences were 100 for 3a and 2a-c, 99 for 1a, 98 for 5a, and 75 for 1b, showing a high robustness of this genotyping method.

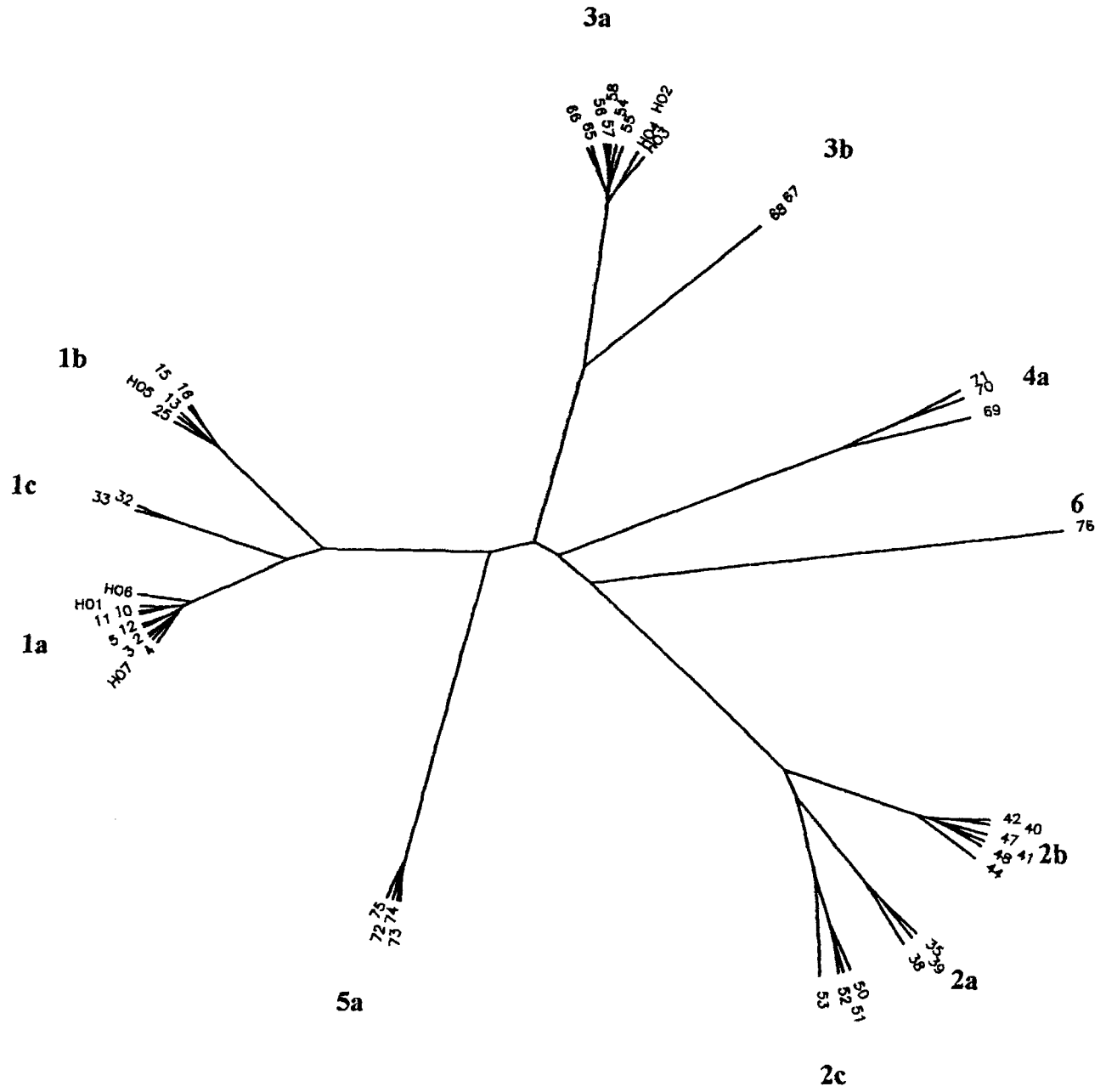

Fig. 1. Phylogenetic analysis of NS5 sequence data. NS5 sequences of 43 HCV strains described earlier [Simmonds et al., 1993] were analysed together with our own data of HCV strains obtained from Honduras. The 43 reference strains are numbered as earlier described, and the Honduran samples are labelled HO 1–7.

Comparison of NS5 and Core Genotyping

The partial NS5 and core regions were amplified and sequenced from 24 European HCV strains using the above-described primers and procedures. Accordant genotyping results were obtained in all of the patients when phylogenetic analyses of the partial NS5 and core regions were performed. Thus, types 1a, 1b, 2b, and 3a were present in those European strains. When the Honduran HCV strains were analysed, the same genotypes were also found as for the NS5 sequences in four strains (HO 1, 6, and 7: type 1a; HO 5: type 1b). A further two HCV strains, HO 3 and 4, were classified as type 1a by phylogenetic analysis of the core sequences (Fig. 2). However, these strains were typed as 3a when the NS5 sequences were used. The sequencing of the core and NS5 regions from these two strains were repeated. No polymorphism, which could indicate a double infection of genotypes 1a and 3a, was identified in the NS5 or the core sequences. In order to analyse further the possibility of a mixed infection, the six Honduran samples were genotyped by a core type-specific PCR, and no type 3a was detected [Okamoto et al., 1993].

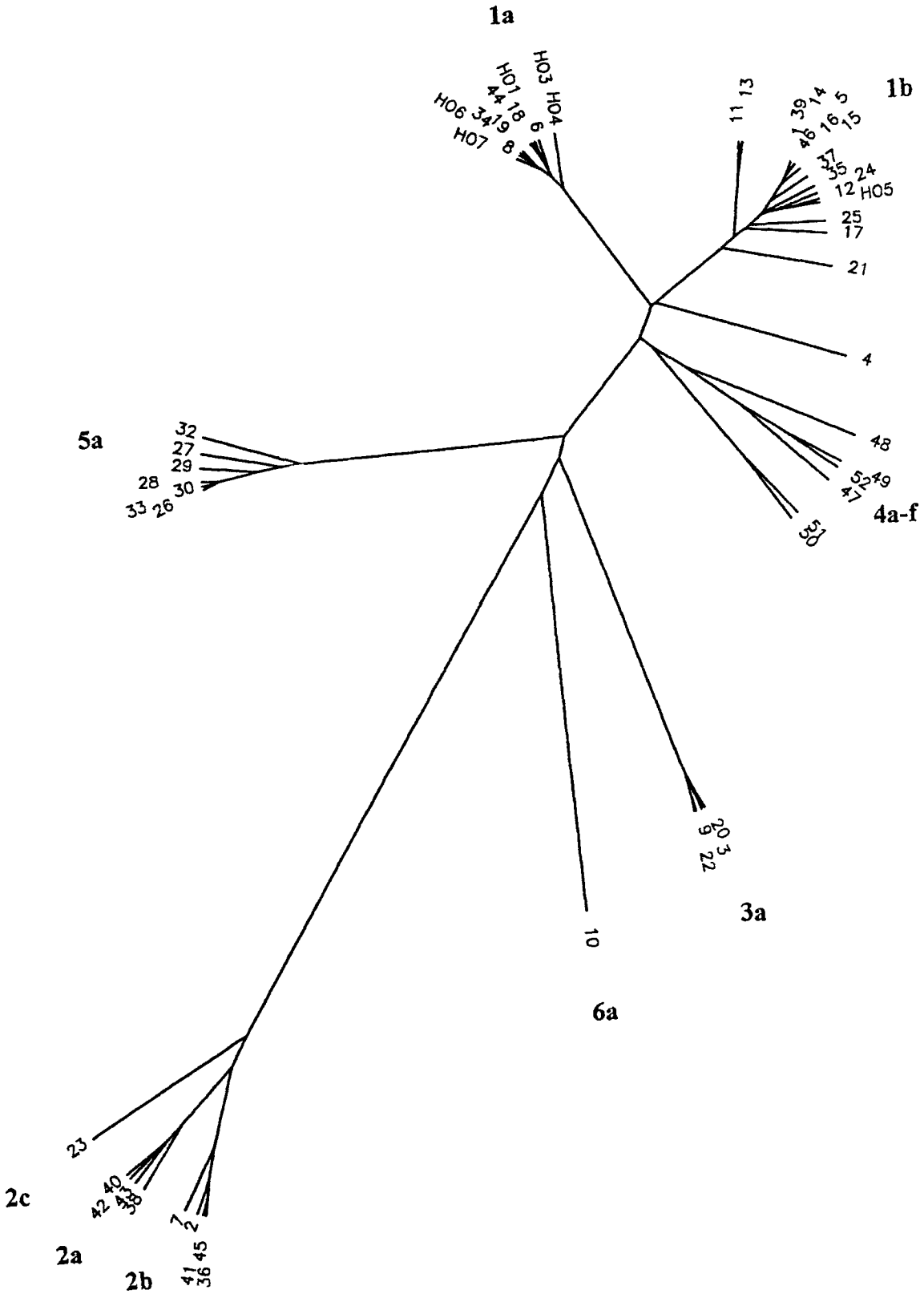

Fig. 2. Phylogenetic analysis of core sequence data. Core sequence of 52 HCV strains described earlier [Bukh et al., 1994] were analysed together with our own data of HCV strains obtained from Honduras. The 52 reference strains are numbered 1–52 in the same consecutive order as the GenBank numbers U10189–10240. The Honduran samples are labelled HO 1 and HO 3–7.

Strain[a]	Genotype	E1 region	E2/NS1 region
Ho7	1a	A G V D A	R T H V T G G S A A R T T A G L A G L F N H G S R Q N I Q L I N
Ho1	1a	- - - - -	T - - - S - - A - G H - - L - F - S - - T S - P S - K - - - V -
Ho4	3a/1a[b]	- - - - -	E - - - - - - - - - - H A A S T F T S I - S P - A K - D - - - - -
Ho5	1b	- - - - G	A - R - - - - A - G H - - S - F - - - - S S - P S - K - - - - -
Ho3	3a/1a[c]	- - - - -	E - - - - - - - - - - H A A S T F - - - - S P - A K - D - - - - -
Ho6	1a	- - - - -	T - Y T - - - A T - - - A - - - T R - - S - - P T - - - - - -
HCV-pt	1a	- - - - -	E - - - - - - - - - G H S A S - F V S - L A T - A K - - V - - - -
HCV-J	1b	- - - - G	H - - - - - - - R V - S S - Q S - V S W L S Q - P S - K - - - V -
HCV-J6	2a	- - - - -	Q - - T V - - - T - H N A R T - T - M - S L - A - - K - - - - -
HCV-J8	2b	- - - - -	T - Y S S - Q E - G - - V - - F - - - - T T - A K - - L Y - - -
NZL1	3a	S - - - -	T - - - - - - - T - - H Y - S R I - - - - - M - P Q - K L - - V -

HVR1 (27 amino acids)

Fig. 3. [a] Predicted amino acid sequences of the N-terminal of the E2/NS1 region amplified from six HCV strains obtained from Honduras and from HCV strains corresponding to genotypes 1a, 1b, 2a, 2b, and 3a [Choo et al., 1989; Kato et al., 1990; Okamoto et al., 1991; Okamoto et al., 1992b; Okamoto et al., 1993]; [b, c] These two strains were classified as type 3a based on analysis of NS5 regions, but 1a based on the analysis of core region.

Analysis of E2/NS1 Sequences

The six Honduran samples were positive when the E2/NS1 region was amplified. As expected, the deduced 27 amino acids at the N-terminal of the E2/NS1 region were highly variable among the different strains (Fig. 3). By comparing the two Honduran strains which belonged to type 1a by analysis of the core region and type 3a by analysis of the NS5 region with published strains of genotypes 1, 2, and 3, it was found that they were similar to the genotype 1a representative strain HCV-PT, but differed greatly from type 3a or other strains.

DISCUSSION

We identified HCV genotypes 1a, 1b, and 3a among Honduran HCV strains, suggesting the presence of multiple HCV genotypes in this Central American country. The N-terminal of the E2/NS1 region was highly variable, confirming the findings regarding the hypervariable feature of this region [Kato et al., 1992]. Interestingly, two Honduras strains exhibited different genotypes when partial NS5 and core sequences were analysed, indicating the existence of chimeral HCV strains in Honduras. This result should be taken into account when genotyping is based on a partial sequence from a single HCV genomic region.

Using carefully designed core primers and well-established NS5 primers, the HCV sequences of genotypes 1, 2, and 3 were expected to be readily amplified. The reliability of the genotyping method using phylogenetic analysis of the selected NS5 and core sequences were further verified by analysing reference strains. The fact that the E2/NS1 sequences of the two strains with discrepant genotypes were more similar to type 1a suggested that the sequences located upstream from the hypervariable region may be derived from the same HCV strain. It is noteworthy that an earlier reported patient, who harboured a chimeral HCV strain, was infected with plural HCV variants [Kato et al., 1992]. HCV superinfection of another patient with more than one genotype has also been described [Kao et al., 1993]. Thus, the possibility of recombination between different HCV strains seems to exist.

If a mixed infection exists in a patient, it is possible that the efficiency of the PCR is different with the different HCV genotypes. Therefore, it cannot be excluded that the discrepancy between the genotype results in the two Honduran strains was due to differences in the amplification rather than a true recombination event. In theory, the predominant HCV genotype should be identified by direct sequencing, but also minor HCV strains consisting of more than 10–15% of the virus population [Leitner et al., 1993]. In our two patients, a de-

tailed sequence analysis did not reveal any polymorphism. Furthermore, no double infections of types 1a and 3a were found using a core type-specific PCR. These data suggest that our genotyping results are not likely to be due to mixed infections.

Accordant genotyping results were, however, obtained when 24 European strains were examined by analysing partial NS5 and core sequences. This result not only further proves the reliability of our genotyping method, but is also in line with the finding that recombination of HCV is a rare event [Simmonds et al., 1994].

ACKNOWLEDGMENTS

This study was supported by the Swedish Work Environment Found. Claudia Lara is student within the Karolinska Institute Research and Training programme (KIRT).

REFERENCES

Bukh J, Purcell RH, Miller RH (1993): At least 12 genotypes of hepatitis C virus predicted by sequence analysis of the putative E1 gene of isolates collected worldwide. Proceedings of the National Academy of Sciences of the USA 90:8234–8238.

Bukh J, Purcell RH, Miller R (1994): Sequence analysis of the core gene of 14 hepatitis C virus genotypes. Proceedings of the National Academy of Sciences of the USA 91:8239–8243.

Choo Q-L, Kuo G, Weiner AJ, Overby LR, Bradley DW, Houghton M (1989): Isolation of a cDNA clone derived from a blood borne non-A, non-B viral hepatitis genome. Science 244:359–362.

Choo QL, Richman KH, Han JH, Berger K, Lee C, Dong C, Gallegos C, Coit D, Medina SR, Barr PJ (1991): Genetic organization and diversity of the hepatitis C virus. Proceedings of the National Academy of Sciences of the USA 88(6):2451–2455.

Enomoto N, Takada A, Nakao T, Date T (1990): There are two major types of hepatitis C virus in Japan. Biochem Biophys Res Commun 170(3):1021–1025.

Houghton M, Weiner A, Han J, Kuo G, Choo QL (1991): Molecular biology of the hepatitis C viruses: implications for diagnosis, development and control of viral disease. Hepatology 14(2):381–388.

Kao J, Chen P, Lai M, Chen D (1993): Superinfection of heterogenous hepatitis C virus in a patient with chronic type C hepatitis. Gastroenterology 105:583–587.

Kato N, Hijikata M, Ootsuyama Y, Nakagawa M, Ohkoshi S, Sugimura T, Shimotohno K (1990): Molecular cloning of the human hepatitis C virus genome from Japanese patients with non-A, non-B hepatitis. Proceedings of the National Academy of Sciences of the USA. 87(24):9524–9528.

Kato N, Ootsuyama Y, Tanaka T, Nakagawa M, Nakazawa T, Muraiso K, Ohkoshi S, Hijikata M, Shimotohno K (1992): Marked sequence diversity in the putative envelope proteins of hepatitis-C viruses. Virus Research 22:107–123.

Lau J, Mizokami M, Kolberg J, Davis G, Prescott L, Ohno T, Perrillo R, Lindsay K, Gish R, Qian K (1995): Application of six hepatitis C virus genotyping systems to sera from chronic hepatitis C patients in the United States. Journal of Infectious Diseases 171:281–289.

Leitner T, Halapi E, Scarlatti G, Rossi P, Fenyö E, Albert J, Uhlén M (1993): Analysis of heterogeneous viral populations by direct DNA sequencing. Biotechniques 15:120–127.

McOmish F, Yap P, Dow B, Follett E, Seed C, Keller A, Cobain T, Krusius T, Kolho E, Naukkarinen R (1994): Geographical distribution of hepatitis C virus genotypes in blood donors: an international collaborative survey. Journal of Clinical Microbiology 32(4): 884–892.

Nakao T, Enomoto N, Takada N, Takada A, Date T (1991): Typing of hepatitis C virus genomes by restriction fragment length polymorphism. Journal of General Virology 72:2105–2112.

Okamoto H, Okada S, Sugiyama Y, Kurai K, Iizuka H, Machida A, Miyakawa Y, Mayumi M (1991): Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human carrier: comparison with reported isolates for conserved and divergent regions. Journal of General Virology 72:2697–2704.

Okamoto H, Sugiyama Y, Okada S, Kurai K, Akahane Y, Sugai Y, Tanaka T, Sato K, Tsuda F, Miyakawa Y (1992a): Typing hepatitis-C virus by polymerase chain reaction with type-specific primers—application to clinical surveys and tracing infectious sources. Journal of General Virology 73:673–679.

Okamoto H, Kurai K, Okada SI, Yamamoto K, Lizuka H, Tanaka T, Fukuda S, Tsuda F, Mishiro S (1992b): Full-length sequence of a hepatitis-C virus genome having poor homology to reported isolates—comparative study of four distinct genotypes. Virology 188(1):331–341.

Okamoto H, Tokita H, Sakamoto M, Hirokita M, Kojima M, Iizuka H, Mishiro S (1993): Characterization of the genomic sequence of type V (or 3a) hepatitis C virus isolated and primers for specific detection. Journal of General Virology 74:2385–2390.

Oshima M, Tsuchiya M, Yagasaki M (1991): cDNA clones of Japanese hepatitis C virus genomes derived from a single patient show sequence heterogeneity. Journal of General Virology 72:2805–2809.

Simmonds P (1995): Variability of hepatitis C virus. Hepatology 21(2): 570–582.

Simmonds P, Holmes EC, Cha TA, Chan SW, McOmish F, Irvine B, Beall E, Yap PL, Kolberg J, Urdea MS (1993): Classification of hepatitis C virus into six major genotypes and a series of subtypes by phylogenetic analysis of the NS-5 region. Journal of General Virology 74:2391–2399.

Simmonds P, Smith D, McOmish F, Yap P, Kolberg J, Urdea M, Holms E (1994): Identification of genotypes of hepatitis C virus by sequence comparison in the core, E1 and NS5 region. Journal of General Virology 75:1053–1061.

Stuyver L, Rossau R, Wyseur A, Duhamel M, Vanderborght B, Van HH, Maertens G (1993): Typing of hepatitis C virus isolates and characterization of new subtypes using a line probe assay. Journal of General Virology 74:1093–1102.

Yun Z, Lindh G, Weiland O, Johansson B, Sönnerborg A (1993): Hepatitis C virus (HCV) RNA related to HCV antibody pattern in serum and liver histology in Swedish blood donors. Journal of Medical Virology 1:37–42.